基于机器学习方法的氢原子钟故障诊断研究*

李荣凯[1,2]，李锡瑞[1]，蔡勇[1]
(1. 中国科学院上海天文台，上海 200030;
2. 中国科学院大学，北京 100049)

摘 要： 氢原子钟作为一种精密的授时守时仪器，在科学研究与工程应用中发挥着重要作用，但目前我国原子钟还存在设备故障、可靠性差等问题。为了简化技术人员排查流程、提升维修效率，文章提出了一种采用机器学习方法，将氢钟历史运行数据作为训练样本，结合DBSCAN及人工神经网络算法得到氢钟诊断模型，从而大幅简化故障排查过程的方法。实验中将训练好的模型部署到嵌入式设备上，并把实时预判结果给技术人员作为参考，证实了这种方法的可行性及有效性。

关键词： 氢原子钟； 故障诊断； 机器学习； 嵌入式
中图分类号： TB939 **文献标识码：** A **文章编号：**

0 引言

氢原子钟在科研工作中具有重要地位，我国氢钟可以满足很多的应用场景[1]，但仍存在故障频发的问题，其中部分状态或信号明显异常的故障，例如无震荡信号、电源电压异常、电离源异常等可以通过参数直接反映，并重新调整以回归正常运行状态。有些故障很难通过参数直接定位，例如接收机异常、控制面板异常等，但可以结合其它现象进一步分析诊断得出结果。这些在一定程度上反映出可以根据氢钟运行的状态参数来判断当前设备是否正常[2]。在用户角度，故障发生时往往很难描述清楚现象，工程师需要实地根据经验诊断，有较大的滞后性，也对全国范围内氢钟维护修理造成了困难。因此若能通过某种诊断方法得出故障类型[3]，既可以简化故障诊断流程，也使得远程指导用户排查维修变得可能。

本文提出了一种基于机器学习算法，使用氢钟历史数据作为样本训练数学模型，而后部署到嵌入式设备之上以实现实时故障诊断的方法。通过比对嵌入式设备和技术人员的诊断结果，证实了这一方法的有效性。此方法主要是根据氢原子钟的物理参数，用机器学习算法判断其运行状态，流程如图所示：

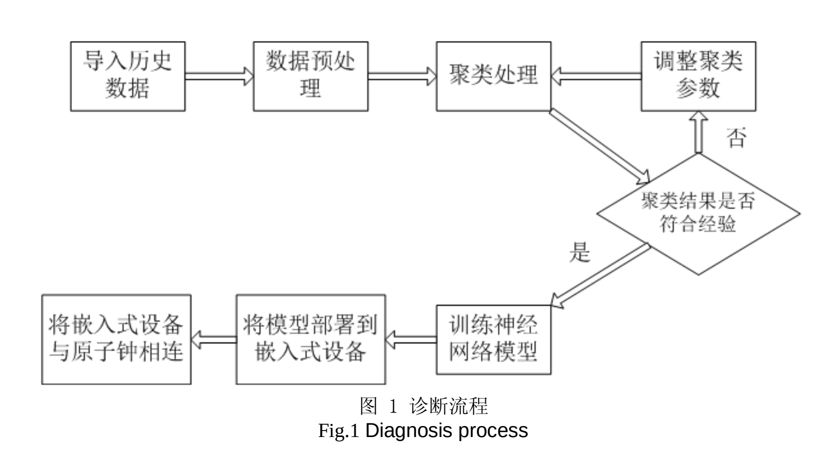


图 1 诊断流程
Fig.1 Diagnosis process

*基金项目：国家自然科学基金资助项目(U1731130)资助.
收稿日期：； 修订日期：
作者简介：李荣凯，男，硕士，研究方向：氢原子钟的智能化研究。 Email:lrk@shao.ac.cn
通讯作者：李锡瑞，男，高级工程师，研究方向：氢原子钟智能控制。 Email:xrli@shao.ac.cn

流程概述如下：

1）调取氢钟运行的历史数据，选择其中7个主要数据作为特征变量，以此导入训练样本。

2）对样本数据进行预处理。

3）使用聚类算法处理数据，得到状态标签，并根据物理意义和诊断经验判断得出的标签是否合适，继而重复调整参数，直至其聚类成带有合适状态标签的数据样本。

4）构建神经网络模型，根据情况设定参数并使用样本数据对模型进行训练。

5）将训练好的算法模型导出，部署到嵌入式设备中。

6）通过串口将嵌入式设备与氢钟相连，进行实时诊断。

综上所述，除一些辅助措施外主要包括聚类算法和神经网络分类两大部分，先通过聚类得到氢钟的运行状态标签，再使用神经网络分类器根据实时数据将当前运行情况归入不同的状态标签中。

1 样本数据处理

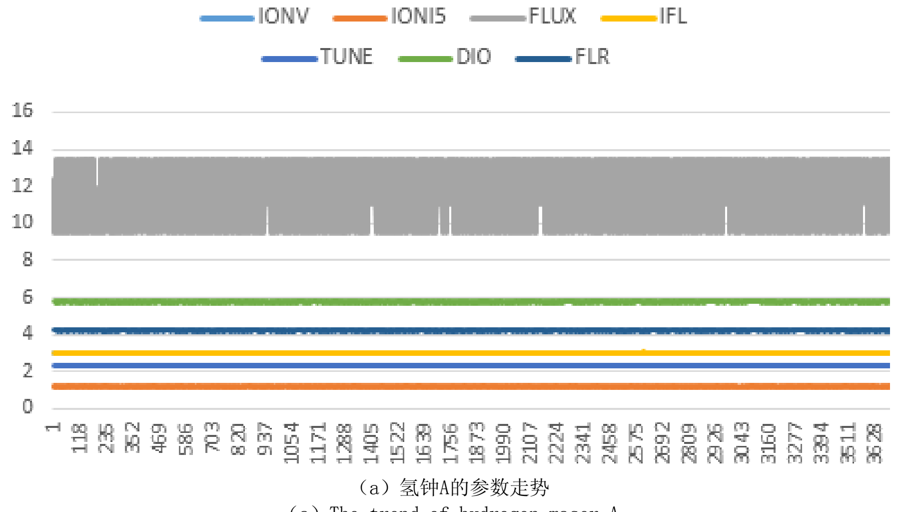
1.1 导入历史数据

目前氢钟已经有了比较完备的监测系统，其中常用的参数有23个。根据维修记录，大部分参数无异常现象，因此决定作降维处理。实验选定了1000组带有人工判断结果的样本，在所得结果与原样本保持高度一致的前提下经过降维将特征数简化为7个，结果降维后的特征与原样本之中的7个参数相似度很高。

为证实这一现象并非偶然，将此7个参数作为特征变量导入样本，然后使用这些样本作聚类处理得到状态标签，并与原样本进行多次比对，得出的结果非常相似。最终结合维修经验及人工判断误差、算法占用芯片资源等因素，直接将这7个参数作为实验特征变量。

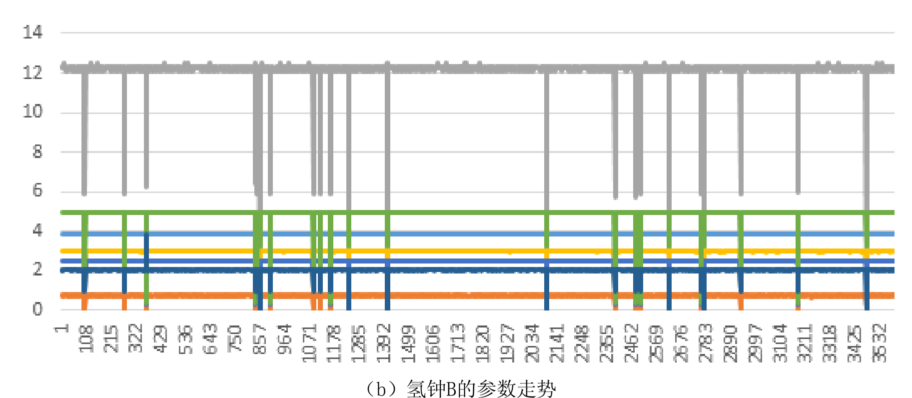
这些参数分别是：离子泵电压(IONV)，离子泵电流(IONI5)，流量(FLUX)，接收机中频电压(IFL)，调谐电压(TUNE)，变容二极管(DIO)和流量设定(FLR)。

不同氢原子钟之间的参数走势差异较大，以三台氢钟参数走势为例：



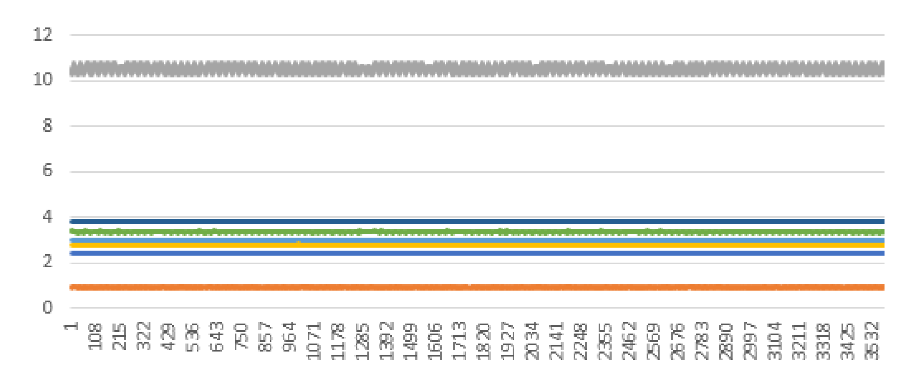
(a) 氢钟A的参数走势

(a) The trend of hydrogen maser A



（b）氢钟B的参数走势

（b）The trend of hydrogen maser B



（c）氢钟C的参数走势

（c）The trend of hydrogen maser C

图 2 三台不同氢钟的参数走势图

Fig.2 data of 3 different devices

从上图可知不同设备即使都运行在正常状态下，实时参数也存在较大差异。对于这种现象，实验中采用聚类算法对特定一台氢钟进行聚类并贴上状态标签，然后这些标签仅能为此台特定氢钟所使用，不为其赋予通用性。

1.2 数据预处理

在训练前对样本数据进行预先处理有时会使所得结果更准确，常用的预处理方法有StandardScaler、MinMaxScaler、RobustScaler、Normalizer等。预处理对氢钟状态预测结果会产生很大影响。

为了确定是否使用预处理，以及使用哪种预处理方式，实验中选取了20000组样本数据，设置不同聚类算法参数(eps和min_samples)，然后观察最终预测准确率，得到如下结果：

表 1 eps为0.8和min_samples为10时的准确率

Tab.1 Predictive accuracy of eps=0.8 & min_samples=10

预处理方法	预测准确率
None	17.76%
StandardScaler	99.42%
MinMaxScaler	95.54%
RobustScaler	96.00%
Normalizer	19.16%

表 2 eps为0.9和min_samples为15时的准确率

Tab.2 Predictive accuracy of eps=0.9 & min_samples=15

预处理方法	预测准确率
None	67.18%
StandardScaler	99.78%
MinMaxScaler	96.30%
RobustScaler	96.44%
Normalizer	66.20%

表 3 eps为1.0和min_samples为20时的准确率

Tab.3 Predictive accuracy of eps=1.0 & min_samples=20

预处理方法	预测准确率
None	96.20%
StandardScaler	99.54%
MinMaxScaler	98.66%
RobustScaler	96.36%
Normalizer	96.18%

可以看出，StandardScaler、MinMaxScaler和RobustScaler三种算法都有不错的效果，其中StandardScaler算法更为合适。

2 聚类得到状态标签

2.1 聚类算法优点

聚类算法可以找出样本之间的差异，自动地将其划分为若干个类别标签，可以避免对庞大的参数样本进行手动状态分类、消耗大量人力物力且容易出差错的问题。事实上，在实验中使用聚类算法将数据分类时出现了很多之前忽略的状态，从而发现了新的故障类型，这也是采用这一方法的重要意义。

常见的聚类方法有K-means算法、ward-linkage算法及DBSCAN算法等，其中常用作故障诊断的方法有K-means和DBSCAN两种算法[4]，实验选择了DBSCAN算法。

2.2 DBSCAN算法

DBSCAN算法又叫作“基于密度的聚类”方法，即根据样本分布的紧密程度来划分样本的类别，并基于某种关系不断扩展聚类簇以获得最终的聚类结果。若将状态标签表示为C，样本集表示为D，则标签划分存在关系[5]：

$$C = \{C_1, C_2, C_3, \ldots, C_k\} \tag{1}$$

样本数据点x_i存在关系：

$$D = \{x_1, x_2, x_3, \ldots, x_m\} \tag{2}$$

将样本点聚为相同类别的依据是满足下述条件的非空子集：

$$x_i \in C,\ \ x_j \in C \Rightarrow x_i\text{与}x_j\text{密度相连}(i, j \in [1, k],\ \ i \neq j) \tag{3}$$

$$x_i \in C,\ \ x_j\text{由}x_i\text{密度可达} \Rightarrow x_j \in C \tag{4}$$

为了找出最合适的聚类算法并对比说明，用一组相同的二维样本点、用两种算法进行聚类得到效果图。

首先使用K-means算法：

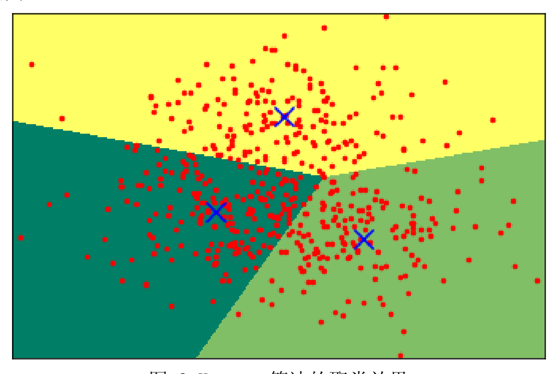


图 3 K-means算法的聚类效果
Fig.3 Clustering result of K-means

然后使用DBSCAN算法：

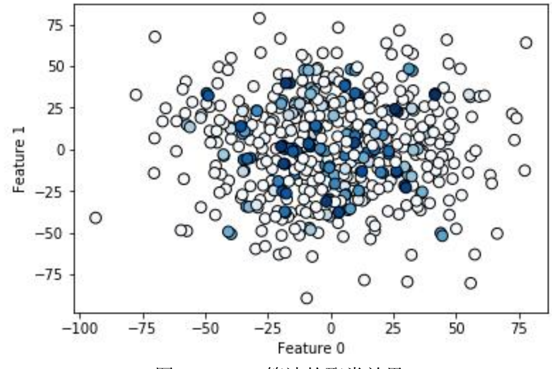
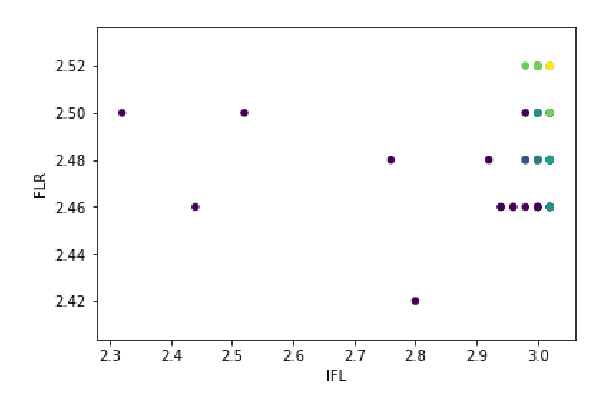


图 4 DBSCAN算法的聚类效果
Fig.4 Clustering result of DBSCAN

图3是以背景颜色划分种类，而图4是以样本点颜色来划分。可以看出K-means算法适合处理相同类别分布比较集中的数据，而DBSCAN算法适合处理不同类别混杂在一起的数据。

若选择FLR和IFL两个参数，不同点的颜色代表不同的状态类别，可以得到氢钟数据的分布图：



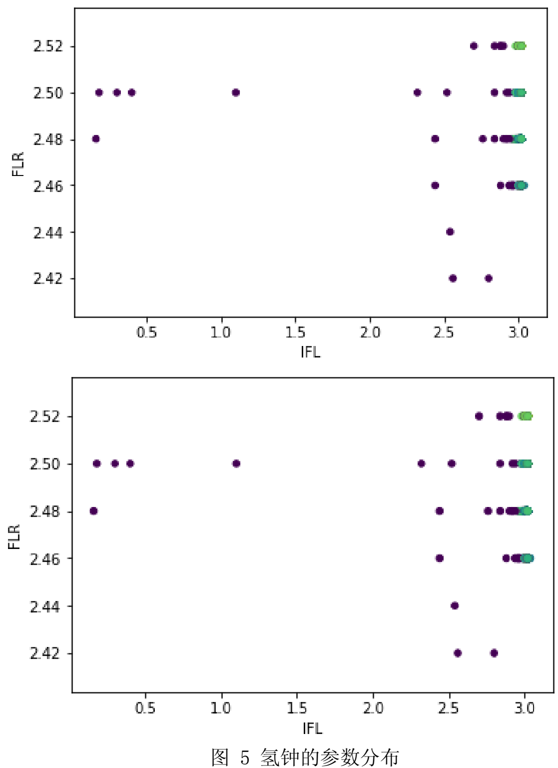


图 5 氢钟的参数分布
Fig.5 Parameter distribution of hydrogen maser

上图分别是同一台氢钟100、1000、10000个样本点，带有18种颜色状态的参数分布图，随着样本增多，很多不同颜色的点混杂在一起，甚至出现肉眼看起来10000个点只能呈现出几十个点的现象。对于样本混杂现象严重的氢钟来说，显然DBSCAN算法更合适。而且根据实验结果，对数据进行适当预处理后DBSCAN算法的表现会更好。DBSCAN算法的另一特点是在聚类时无需指定要分成的类别数量，这对状态判断困难的氢钟来说更具灵活性。

2.3 对标签进行调整

实验中发现，有时聚类得到的状态标签并不适合用来直接训练分类模型。

使用聚类算法时需要设定大致要分成的状态数量，根据以往经验一台氢钟常见的工作状态有7到9种，然而之后发现这种设定方式会产生较大的误判率。比如在测试某台钟时，其正常运行时的平均参数值比历史数据的峰值小很多，从而出现了所有参数变为0时，屏幕上显示的判断依然为“正常”的现象。因此直接分成7到9种状态有时并不符合实际的应用情况，也往往难以匹配合适的物理意义。

在多次试验总结后，最终采用了先将聚类标签数量设定为15到90种，然后根据每种状态的物理意义对标签进行二次调整的方法：

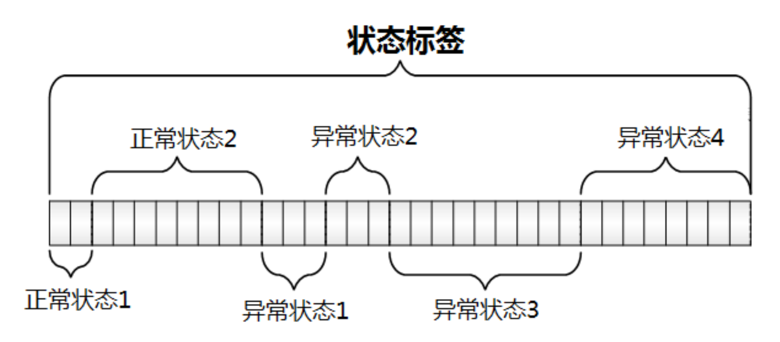


图 6 对标签进行二次分类

Fig.6 Secondary classification of cluster labels

3 根据状态标签进行分类

3.1 引入分类器

数据经过聚类处理后可以得到氢钟常见的状态标签，工程师们可以根据每一种或数种状态的典型物理特征将状态贴上标签，如：正常状态、电离泡亮度异常、接收机失锁等等。

此时，若把打上标签的历史数据作为样本来训练分类模型，将其中的状态参数作为特征变量，将标签作为目标变量，我们便可以通过训练好的分类模型实时处理数据，从而达到状态检测、故障判定的效果。本文选择了人工神经网络作为分类算法，下文将对此作出解释。

3.2 选择分类算法

常见的分类算法有：决策树(Decision Trees)、人工神经网络(Artificial Neural Network, ANN)、K最近邻算法(k-Nearest Neighbor, KNN)、支持向量机(Support Vector Machine, SVM)、朴素贝叶斯(Naive Bayes model)等[6,7,8]。它们在实际应用中各有优缺点，比如决策树算法便于观察和理解，但处理存在缺失的样本效果不好，而且容易丢失特征变量间的相关性；KNN算法方便快捷，擅于处理内容重复较多的样本，不过学习速度较慢，且类别数量过多时容易出问题；支持向量机算法擅长处理非线性、高维问题，并能防止局部极小点情况出现，但对数据完整性要求较高，且不便于处理非线性问题。[9]

ANN算法的优点是分类准确度高，且有不错的鲁棒性；缺点是对于结果难以解释，属于黑盒模型。在氢钟状态分类问题上更看重准确性和鲁棒性，对于结果是否可解释并不关心。[10]

从氢钟的历史数据上看，即使是正常运行的设备，其参数也可能相对不稳定：

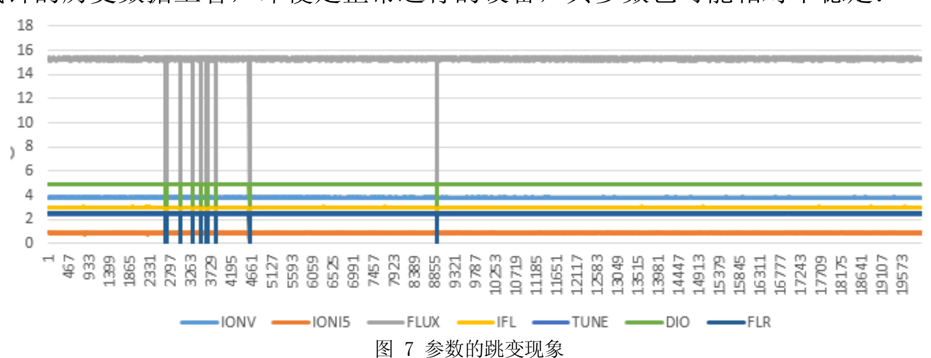


图 7 参数的跳变现象

Fig.7 Parameter jump phenomenon

从上图可以看出，设备存在跳变现象，对于算法的鲁棒性要求较高。故此实验中选择了人

工神经网络作为分类算法。

3.3 神经网络概述

神经网络算法原理来源于生物的神经系统，其基本的成分是神经元(neuron)模型。若以x为输入，y为输出，ω为第i个神经元的连接权重，f()为激活函数，输入与输出存在关系：

$$y = f\left(\sum_{i=1}^{n} \omega_i x_i - \theta \right) \tag{5}$$

多个神经元按照一定原理相组合，便可以得到人工神经网络。

3.4 不同分类算法结果对比

下表列出了在不同聚类标签数量时决策树、ANN和KNN算法进行分类的准确率：

表 4 不同算法的预判准确率比较

Tab.4 Comparison of accuracy rates with different algorithms

状态标签数量	DecisionTree	ANN	KNN
7	99.42%	99.68%	99.59%
14	99.44%	99.63%	99.54%
18	99.04%	99.48%	99.46%
72	92.95%	99.68%	99.60%

实验中神经网络设置为2层，每层30个节点，激活函数为tanh；决策树的最大深度是5。由表4可以看出对于氢钟状态预测，ANN是比较合适的算法。

4 实验结果

由于机器学习算法占用芯片资源较大，实验中选择搭载A57内核的开发板以满足算力要求。将算法部署到嵌入式设备便于做成手持充电式氢原子钟故障诊断仪，也可以集成入氢钟电路内部。实验中将上位机训练好的模型移植到嵌入式设备上，把串口线一端接电路板，另一端插入氢钟后侧的接口即可[11]，数据便会从氢钟内部读入嵌入式设备，并在屏幕上显示结果，如下图所示：



图 8 结果照片

Fig.8 Result photo

在与原子钟相连后，嵌入式设备屏幕上会循环显示数据，如下图照片所示。其中”Main parameters are”一行显示氢钟当前运行的参数数值,”Now the hydrogen maser working in state”一行显示状态标签代码：

```
Main parameters are:   IONV:  3.84 IONI5:  0.26 FLUX:  0.0
Now the hydrogen maser working in state: [1]

Main parameters are:   IONV:  3.84 IONI5:  0.28 FLUX:  0.0
Now the hydrogen maser working in state: [1]

Main parameters are:   IONV:  3.84 IONI5:  0.28 FLUX:  0.0
Now the hydrogen maser working in state: [1]

Main parameters are:   IONV:  3.84 IONI5:  0.28 FLUX:  0.0
Now the hydrogen maser working in state: [1]

Main parameters are:   IONV:  3.84 IONI5:  0.28 FLUX:  0.0
Now the hydrogen maser working in state: [1]

Main parameters are:   IONV:  3.84 IONI5:  0.26 FLUX:  0.0
Now the hydrogen maser working in state: [1]
```

图 9 嵌入式屏幕照片
Fig.9 Embedded device screen photo

5 结语

通过状态参数判断氢原子钟运行状态是工程师们常用的方法，因此氢钟历史数据对于故障诊断具有一定的参考价值。文中阐述的基于此的机器学习方法具有较强的实用性，算法模型也可以通过重复训练增强通用性。实验中也可以设定程序，使氢钟根据新产生的数据周期性迭代训练模型，以保持算法模型的有效性。

目前仅有部分氢钟的数据可用，在其中更是鲜有带有频率输出数据的样本，因此随着氢钟样本不断增多和完善，文章中提出的方法还有很大提升空间。综上，使用机器学习方法使原子钟具备自诊断，甚至自修复功能的思路，对于氢钟可靠性提升来说是一个很好的方向。

Research on Fault Diagnosis of Hydrogen Maser Based on Machine Learning

Rongkai Li[1,2], Xirui Li[1], Yong Cai[1]
(1. Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Hydrogen maser, as a precise time-keeping and punctual instrument, plays an important role in scientific research and engineering application. However, at present, hydrogen maser in China still have problems such as equipment failure and poor reliability. In order to simplify the troubleshooting process for technicians and improve maintenance efficiency, This article proposes a method that uses machine learning to take hydrogen maser historical data as training samples, combines DBSCAN and artificial neural network algorithms to obtain a clock diagnostic model, thereby greatly simplifying the troubleshooting process. The trained model was deployed on the embedded system in the experiment, then real-time prediction result was given to the technical staff as a reference, which confirmed the feasibility and effectiveness of this method.

Key words: Hydrogen maser; Fault diagnosis; Machine learning; Embedded system

参考文献:

［1］ 王国瑞,杜燕,李锡瑞.氢脉泽调制技术和研究[J].天文研究与技术,2018,15(04):473-478.

Wang Guorui, Du Yan, Li Xirui.Hydrogen maser modulation technology and research [J] .Astronomical Research and Technology, 2018,15 (04): 473-478

［2］ 邓小波,赵斌,王铮.原子钟在线监测评估方法设计[J].无线电工程,2018,48(01):46-49

Deng Xiaobo,Zhao Bin,Wang Wei.Design of online monitoring and evaluation method for atomic clock[J].Radio Engineering,2018,48(01):46-49

［3］ Ye F, Zhang Z, Chakrabarty K, et al. Board-Level Functional Fault Diagnosis Using Multikernel Support Vector Machines and Incremental Learning[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2014, 33(2):279-290

［4］ 肖黎,罗嘉,欧阳春明.基于半监督学习方法的磨煤机故障预警[J].热力发电,2019,48(04):121-127

Xiao Li,Luo Jia,Ouyang Chunming.Fault warning of coal mill based on semi-supervised learning method[J].Thermal Power Generation,2019,48(04):121-127

［5］ 周志华,机器学习[M].清华大学出版社,2016(1):210-214

Zhou Zhihua,Machine Learning,[M].Tsinghua University Press,2016(1):210-214

［6］ 周爱国,王嘉立,杨思静,沈勇,楼狄明.基于K-means和K近邻的DPF设备故障分类算法[J].内燃机与配件,2019(12):57-59

Zhou Aiguo, Wang Jiali, Yang Sijing, Shen Yong, Lou Diming.A Fault Classification Algorithm for DPF Equipment Based on K-means and K Near Neighbors[J].Internal Combustion Engines & Parts,2019(12):57-59

［7］ 董嘉真.人工神经网络在分类问题中的应用[J].科技传播,2019,11(02):143-144

Dong Jiazhen.Application of Artificial Neural Network in Classification Problem[J].Science and Technology Communication,2019,11(02):143-144

［8］ Aykut Erdamar,Mehmet Feyzi Aksahin. Multi-scale classification of single-cell gel electrophoresis assay using deep learning algorithm[J]. Biomedical Signal Processing and Control,2019,56

［9］ 朱梦. 基于机器学习的中文文本分类算法的研究与实现[D].北京邮电大学,2019

Zhu Meng. Research and Implementation of Chinese Text Classification Algorithm Based on Machine Learning [D]. Beijing University of Posts and Telecommunications, 2019

［10］ 王子玲,贾舒宜,修建娟,陈小慧.基于人工神经网络的多模型目标跟踪算法[J].海军航空工程学院学报,2019,34(04):343-348+370

Wang Ziling,Jia Shuyi,Jian Juan,Chen Xiaohui.Multi-model target tracking algorithm based on artificial neural network[J].Journal of Naval Aeronautical Engineering Institute,2019,34(04):343-348+370

［11］徐艺灵,张鸿飞,姜逢欣,陈金挺,陈亚奇,贾明皓,董书成,陈杰,杨臣威,王坚.基于STM32的可调恒温滤光片转轮盒的设计[J].天文研究与技术,2018,15(03):323-331

Xu Yiling, Zhang Hongfei, Jiang Fengxin, Chen Jinting, Chen Yaqi, Jia Minghao, Dong Shucheng, Chen Jie, Yang Chenwei, Wang Jian. Design of STM32-based adjustable constant temperature light film runner box [J]. Astronomical Research and Technology, 2018, 15 (03)): 323-331